

全国棉花种植面积遥感监测抽样方法设计

焦险峰, 杨邦杰, 裴志远

(中国农业工程研究设计院)

摘要: 棉花是我国重要的经济作物之一, 准确、及时地提供棉花的播种面积, 对政府部门及时掌握棉花生产情况, 制定相关调控政策, 对棉花加工企业与棉农预测市场情况有重要意义。由于中巴卫星的应用, 对新疆农情的及时遥感监测成为可能, 从而使遥感监测可以覆盖全国。但是, 考虑到中国东、中、西部县级面积相差很大, 必须设计不同的抽样方法。在新疆棉花监测的抽样设计中, 采用了标准地形图幅作为分层抽样的抽样单元的方法。在黄淮海、长江中下游棉区采用县为单元的分层抽样的方法, 抽样样本用遥感调查的方法获得。设计了满足生产应用的全国作物面积遥感监测抽样运行方法。

关键词: 棉花; 种植面积; 遥感监测; 抽样调查

中图分类号: S127

文献标识码: A

文章编号: 100226819(2002)0420159204

棉花是我国重要的经济作物之一。自 1950 年以来, 一直是政府制定棉花价格并由供销社统一经营。20 世纪 50 年代初, 全国棉花播种面积 378.6 万 hm^2 。60、70 年代我国棉花面积平均 467 万 hm^2 , 进入 90 年代后, 中国的棉花价格主要由市场形成。棉花面积受价格的影响逐年波动, 棉花播种面积最高年份达 684 万 hm^2 (1992 年), 一般年份面积平均 520 万 hm^2 。经过多年发展, 目前形成黄淮海流域、长江流域和新疆三大产棉区。

准确、及时地提供棉花的播种面积和产量, 对政府部门及时掌握棉花生产情况制定相关调控政策, 对棉花加工企业与棉农预测市场情况有重要意义。农业部农业资源监测站自 1999 年起应用美国陆地卫星遥感影像监测棉花播种面积。到 2001 年连续 3 年监测黄淮海和长江流域两大棉区的棉花面积变化。2000 年中巴资源卫星的应用, 为遥感监测新疆棉区的棉花面积提供了可能。2001 年监测了新疆棉区棉花面积变化, 从而首次用遥感手段监测到全国棉田的变化。本项研究的目的是设计实用的棉花种植面积遥感监测方法, 关键是抽样方法。

1 作物面积监测的基本思路

农作物种植面积的大范围遥感监测一般采用抽样方法。欧盟 MARS 计划、美国大面积农作物估产计划、农业和资源的空间遥感调查计划等都应用了抽样方法。中科院在全国农情遥感监测与估产运

行化方法中提出在高精度耕地数据库的支持下, 通过提取不同作物种植成数估算作物种植面积的抽样方法。农业部门提出大尺度的作物种植面积变化遥感抽样方法, 得到能够满足生产应用的实际监测运行方法。

本项研究采用分层抽样的方法设计总体技术模型, 抽样样本用遥感调查的方法获得, 推算棉花面积年度间变化, 从而实现棉花种植面积的遥感监测。中国东部、中部与西部县域的面积和种植制度等情况相差很大, 需要在不同地区设计不同的抽样框架。2001 年全国范围的棉花面积遥感监测在黄淮海、长江中下游棉区和新疆棉区分别进行。

2 新疆棉区遥感监测抽样方案的设计

2.1 抽样框的确定

抽样单元的构造考虑两种原则: 满足组织实施的便利; 尽可能提高抽样精度, 减少估计量的方差。

新疆共 85 个县市, 县域面积很大, 不可能选择县为抽样单元。为保证精度, 必须增加样本数。因此, 选择标准地形图幅为抽样单元。通过试算, 选择 1:25 万比例尺的地形图为抽样单元比较合适。新疆整个疆域面积内, 共有 1:25 万比例尺地形图 18 628 个, 有耕地分布的单元 2 831 个, 其中分布在有种植棉花记录的县域内的单元 2 080 个。新疆棉花遥感面积监测分层抽样就以这 2 080 个单元为抽样框。

2.2 分层标志的选择

2.2.1 选择分层标志的依据

要选择恰当的变量作为分层标志, 最好采用已有的调查数值, 但调查尚未进行, 这些数值还是未知的。通常选择一个与调查指标有较大线性相关的指

收稿日期: 200203211

基金项目: 国防科工委中巴卫星应用示范项目; 农业部发展计划司棉花监测项目

作者简介: 焦险峰, 工学硕士, 高级工程师, 北京朝阳区麦子店街 41 号 中国农业工程研究设计院农业部农业资源监测总站, 100026

标作为分层标志。基于地区种植制度和农民种植习惯,历年播种面积与调查指标—当年播种面积的相关程度高。在棉花播种面积调查中选择上一年棉花播种面积作为分层的指标。

2.2.2 分层标志的计算

以地形图为抽样单元,没有现成的以地形图为单位的统计数据。本次又是第一次用遥感手段调查新疆棉花面积,没有现成的棉花面积分布图。必须设计一种算法将统计数据折算到地形图图幅范围内,得到以地形图图幅范围为单位的数据。其算法是:假设各县和新疆兵团的棉花种植面积均匀分布于各行政区域内的耕地上。计算抽样单元内的耕地面积占所在行政区域耕地面积的百分比,以此比例为权重,将各行政区域棉花播种面积分解到各个抽样单元,得到各抽样单元的棉花播种面积。

新疆棉花播种面积是按照县级行政区和建设兵团 2 个途径统计的,收集这 2 个途径的统计数据的同时收集了新疆行政区域分布图和建设兵团分布图,在 GIS 的支持下,将这两张分布图叠加成一张新疆行政区和建设兵团分布图。在 1:100 000 全国资源环境数据库中的耕地层数据的基础上,分别计算各行政区的耕地和抽样单元内的耕地面积,最后,结合收集到的各行政区域统计数据,计算得到各抽样单元的棉花播种面积。

2.3 分层

2.3.1 层数的确定

一般地说分层数目越多,抽样统计越准确,但抽样方差的降低是与分层数的平方成反比。分层抽样理论要求每层至少必须抽取 2 个样本单元,层数不超过 $n/2$ 个。但是在层数大于 6 时方差的减少幅度将大为减缓。理论与实践研究表明层数为 6 时,能够满足抽样要求。

2.3.2 累计平方根法分层

用累计平方根法确定各层的分点。最优抽样要使所抽取数值的层方差和最小。根据 y 的密度函数 $f(y)$ 的开方累计规划计算分层。将抽样单元的棉花播种面积以 100 hm^2 的组距分组,计算各组频率、频率累加值和频率开方累加,得到分层界线。各层分层界线见表 1。最终将所有抽样单元分为 6 层。

表 1 分层参数表

Table 1 The variable values in each stratum

层	分层界线	样本数	层总值
1	87.2	1 245	35 959.4
2	174.4	194	74 359.1
3	261.6	283	123 847.5
4	348.8	186	169 085.8
5	436.0	117	218 041.5
6	523.2	55	265 832.4

2.3.3 分层图的制作

在 GIS 的支持下,将分层结果制作成分层图,以掌握样本在各层的分布情况,为下一步遥感方法获取抽样样本过程中遥感数据的选择提供依据。

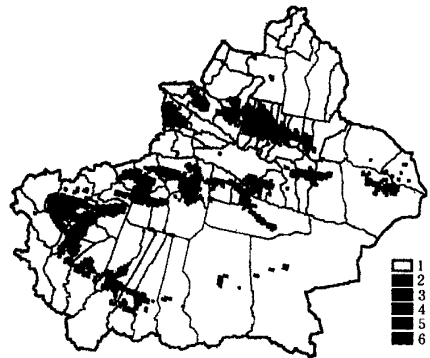


图 1 新疆以 1:25 000 比例尺地形图为抽样单元的分层抽样分层图

Fig. 1 Distribution of stratum with a sampling unit of 1:25 000 standard map sheet

2.4 样本数量的估算

比例分配分层抽样需要的最小样本数 n 由下式决定

$$n_0 = \frac{N \sum_{h=1}^L N_h s_h^2}{V}; n = \frac{n_0}{1 + \frac{n_0}{N}}$$

$$\text{比例分配抽样比 } f = \frac{n}{N} = \frac{n_h}{N_h}$$

式中 V —— 估计量的方差,一般给定绝对误差界限 d , 则

$$V = \left(\frac{d}{u_{\frac{\Delta}{2}}}\right)^2$$

若精度的要求以相对误差限 A 的形式提出, 则 $u_{\frac{\Delta}{2}}$ 是正态分布的上 $\frac{A}{2}$ 分位数;

L —— 分层抽样的层数; N_h —— 各抽样层样本数目, $h = 1, 2, \dots, L$; n_h —— 各抽样层抽取用于棉花面积调查的样本数, s_h 层的样本方差

$$s_h = \frac{1}{n_h - 1} \sum_{i=1}^{n_h} (y_{hi} - \bar{y}_h)^2$$

在抽样精度 95% 的情况下,计算出新疆棉花遥感面积监测分层抽样的样本数为: $n = 204$, 即 204 个 1:2.5 万比例尺的地形图,为 2 080 个地形图的 9.8%。各层都抽取该层总数目的 9.8%, 各层抽样数见表 2。

表 2 抽样结果

Table 2 The variable values for each stratum

层	总样本数	抽样比δ%	抽样数	层权重
1	1 245	9.8	122	0.598
2	194	9.8	19	0.093
3	283	9.8	28	0.136
4	186	9.8	18	0.089
5	117	9.8	12	0.056
6	55	9.8	5	0.026
总计	2 080		204	

2.5 外推方法对总量的估计

通过比例分配分层抽样后, 播种面积的总体均值的估计值 \bar{Y}_{st} 为

$$\bar{Y}_{st} = \frac{1}{N} \sum_{h=1}^L N_h \bar{Y}_h = \sum_{h=1}^L W_h \bar{y}_h = \sum_{h=1}^L W_h \frac{1}{n_h} \sum_{i=1}^{n_h} y_{hi}$$

通常情况下,

$$\bar{y}_{st} = \bar{y}, y = \frac{1}{n} \sum_{h=1}^L n_h \bar{y}_h$$

\bar{y}_{st} 方差的无偏估计量为

$$V_{st}(\bar{y}) = s^2(\bar{y}) = \sum_{h=1}^L \frac{W_h^2 s_h^2}{n_h} - \sum_{h=1}^L \frac{W_h s_h^2}{N} = \frac{1-f}{n} \sum_{h=1}^L W_h s_h^2$$

$$s(\bar{y}_{st}) = \sqrt{V_{st}(\bar{y})}$$

总体均值和总体总值的置信区间分别为 $\bar{y}_{st} \pm u_{\frac{\alpha}{2}} s(\bar{y}_{st})$ 和 $N \bar{y}_{st} \pm u_{\frac{\alpha}{2}} N s(\bar{y}_{st})$ 。

假设 \bar{y}_{st} 服从正态分布, $u_{\frac{\alpha}{2}}$ 为正态分布上侧 $\frac{\alpha}{2}$ 分位数, $1 - \alpha$ 为置信度。

上面各式中 L —— 分层抽样的层数; N_h —— 各抽样层样本数目, $h = 1, 2, \dots, L$; n_h —— 各抽样层抽取用于棉花面积调查的样本数; Y_{hi} —— 第 h 层第 i 个样本县棉花播种面积; W_h —— 层权, $W_h = \frac{N_h}{N}$ 。

$$h \text{ 层的总体均值: } \bar{Y}_h = \frac{1}{N_h} \sum_{i=1}^{N_h} Y_{hi}$$

$$h \text{ 层的总体方差: } S_h = \frac{1}{N_h - 1} \sum_{i=1}^{N_h} (Y_{hi} - \bar{Y}_h)^2$$

$$h \text{ 层的样本均值: } \bar{y}_h = \frac{1}{n_h} \sum_{i=1}^{n_h} y_{hi}$$

$$h \text{ 层的样本方差: } s_h = \frac{1}{n_h - 1} \sum_{i=1}^{n_h} (y_{hi} - \bar{y}_h)^2$$

2.6 计算

在 2001 年全国棉花播种面积遥感监测业务化

运行中, 应用本方法监测新疆棉区棉花播种面积的变化, 订购同一地区 2000 年和 2001 年两年的中巴资源卫星数据 42 景, 图象解译后, 将解译结果按照地形图图幅统计, 得到抽样样本, 进一步用外推方法估算总量和变化量。在 95% 的抽样精度和 95% 的置信度下, 得到 2001 年度新疆棉区棉花播种面积的变化。计算结果精度达到业务化运行要求。

3 黄淮海和长江中下游棉区抽样方案设计

黄淮海和长江中下游棉区是我国传统的棉区, 棉花种植多以家庭分散型为主, 棉田分布多是零星地块, 在分层抽样中, 考虑抽样实施和样本数据获取的方便, 确定以县为抽样基本单元。按照中国棉花种植区划, 全国棉花生产区共有 1 000 多个县(区), 但其中有些县市棉花播种面积多年平均面积为零, 实际全国棉花种植县为 900 多个, 去掉新疆棉区, 黄淮海和长江中下游棉区共有 875 个县。以此为黄淮海和长江中下游棉区棉花种植面积抽样调查的总体。

与新疆棉区遥感调查方法相同, 采用遥感调查和分层抽样相结合的方法估算 2001 年黄淮海和长江中下游棉区棉花播种面积变化。以该区范围内的 875 个县为抽样单元构建抽样框, 选择 1998 年棉花播种面积的统计值为分层指标, 累计平方根法分层。分层后用比例分配的方法计算各层样本数, 计算出抽样的样本数。订购同一地区连续两年的 TM 卫星数据。将卫星影像覆盖范围对棉花面积的抽样调查近似的看作随机抽样。图像解译后, 将解译结果按照地形图图幅统计, 得到抽样样本。估算抽样区的总体总值和总体均值, 计算两年间播种面积的变化。

通过 2000 和 2001 年连续两年采用上述方法监测黄淮海和长江中下游棉区棉花播种面积变化的结果表明, 在 85% 的抽样精度和 95% 的置信度下, 利用 18 景 TM 数据对该区棉花播种面积变化的监测精度可以达到应用部门业务化运行的要求。

4 抽样方法的实现与运行

为便于业务化运行, 利用组件式地理信息系统平台软件 MAPOBJECT 和 VB 开发了全国棉花种植面积遥感监测管理系统, 编程实现了给定任意参数的分层抽样和外推模型。系统包括 3 个模块: 全国棉花生产背景资料管理; 计算机自动分层抽样与外推; 文档资料的管理。系统主要实现以下功能: 查询显示行政区划、自然资源和各县棉花生产基本情况等背景资料; 计算机自动完成分层抽样, 显示分层结果图; 辅助用户抽样, 并显示每层随机抽样结果; 输入棉花面积遥感解译结果; 利用分层抽样

外推模型计算棉花播种面积及其变化率、置信度和精度; 检索、存储、修改文档文件。

2001 年应用上述抽样理论和“全国棉花种植面积遥感监测管理系统”分别对全国三大棉区的棉花播种面积完成了遥感抽样调查。对全国棉花面积变化的监测精度达到了业务化运行的要求和为宏观决策服务的目标。

5 结 论

本研究设计了可业务化运行的全国棉花种植面积遥感监测抽样方法, 特别是设计了符合新疆棉区特点的以地形图为抽样单元的分层抽样方法, 并编制了“全国棉花种植面积遥感监测管理系统”监测全国棉花播种面积的变化。通过理论研究和 2001 年的业务化运行实践, 得出以下结论:

1) 中巴卫星的应用, 可对新疆农作物的及时遥感监测, 从而使覆盖全国的农作物遥感监测成为可能。但是, 考虑到中国东、中、西部县级面积相差很大, 必须设计不同的抽样方法。

2) 在新疆棉花监测的抽样设计中, 采用了标准地形图幅作为分层抽样的抽样单元的方法。按标准分幅的地形图图幅范围是一定的, 图幅编号是唯一的, 可以方便的列出抽样单元的清单, 构建抽样框; 实际应用中可以选择 1:100 000、1:50 000、

1:25 000 或 1:10 000 各种比例尺的地形图为抽样单元, 满足抽样理论对样本数量的要求。

3) 在 1:100 000 高精度耕地数据库的支持下, 将以县为单位的统计数据折算到某一范围的地理单元的算法基本上能够满足抽样的要求。在 2002 年的新疆棉花遥感面积调查中可以应用 2001 年调查的抽样样本数值作为分层标志数值的补充。

[参 考 文 献]

- [1] L. Kish. 抽样调查[M]. 北京: 中国统计出版社, 1997.
- [2] 刘海启. 大尺度耕地变化监测的遥感抽样方法研究[J]. 农业工程学报, 2001, 17(2): 168~171.
- [3] 冯士雍, 倪加勋, 邹国华. 抽样调查理论与方法[M]. 北京: 中国统计出版社, 1998.
- [4] 吴炳方. 全国农情监测与估产的运行化遥感方法[J]. 地理学报, 2000(1): 23~35.
- [5] ESR I Map objects programmers reference[R]. Redlands, California, USA: Environmental System Research Institute, Inc 1996.
- [6] 中华人民共和国农业部. 中国农业统计资料[M]. 北京: 中国农业出版社, 1998. 27~70.
- [7] 刘纪远. 中国资源环境遥感宏观调查与动态研究[M]. 北京: 中国科学技术出版社, 1997.
- [8] 全国农业区划委员会. 中国农业自然资源和农业区划[M]. 北京: 中国农业出版社, 1998. 78~81.

Design of Sampling Method for Cotton Field Area Estimation Using Remote Sensing at a National Level

Jiao Xianfeng, Yang Bangjie, Pei Zhiyuan

(Chinese Academy of Agricultural Engineering, Beijing 100026, China)

Abstract: Timely and accurate estimations of cotton field area at a national or provincial level are essential to management decisions of farmers and the government. As the China-Brazil Earth Resources Satellite (CBERS21) was successfully launched in 1999, it is possible to monitor crops using remote sensing in the northwestern China's Xinjiang Uygur Autonomous Region and the cotton's annual acreage variation at a national level using remote sensing. A stratified sampling method is used to estimate the cotton field area in China and sampled data are collected by CBERS21 images. Considering area of counties and crop fields is very different between the eastern and western China, different sampling units are used in the estimation. A 1:250 000 standard map sheet (about 91 km²) is used as a sampling unit in the western China's Xinjiang Uygur Autonomous Region and one county is used as a sampling unit in eastern China. The sampling method is operationally used in the crop monitoring system for the Ministry of Agriculture.

Key words: cotton; cultivation area estimation; remote sensing; sampling method